

CHAPTER 2

LITERATURE REVIEW

In this chapter, the study of different types of web sites is presented. Section 2.1 discusses the definition and structure of a typical web site. In Section 2.2, seven different types of web sites are discussed. Starting in Section 2.3, five content-rich web sites are reviewed and analysis is made on each of them. Considerations are made on the way the site present and organize their contents. Section 2.3.1 looks at the Library of Congress web site. Section 2.3.2 presents the National Archives of Australia web site. Section 2.3.3 looks at the American Museum of Natural History web site. Section 2.3.4 considers the National Museum of American Art and finally Section 2.3.5 looks at the country's archives web site: the National Archives of Malaysia web site. Section 2.4 summarizes this chapter by emphasizing the needs for a more structured web site.

2.1 WEB SITE: A DEFINITION AND STRUCTURE

Simply defined, a web site is one or more pages of information on the Internet. It is a group of web pages linked for a common purpose with a common theme [Young, 1999]. Web sites are often contained at a single physical location or server.

What is always in common is the single purpose and theme. Well-formed web sites have all pages accessible by easy navigation techniques. A web site may have a single page or it may have thousands or even millions in the case of content-rich web sites.

The home page and each page within the site contain references (hyperlinks) that will take users directly to other pages for more information on a subject. These interactive links allow the user to click on a word or image to go to another page within the same Web site or the link may connect the user to another Web site anywhere in the world. This layering of information gives the user as much data as they wish to see on a subject.

A web page is normally dynamic. It can be modified easily by the author to contain the latest information, unlike a brochure that may be out of date by the time it is printed. The web site is an electronic publication that gives the world access to information about an individual or an organization on information they wish to share.

A typical web site structure looks like the following Figure 2.1, which is hierarchically structured as a tree.

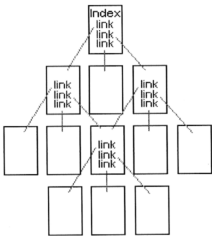


Figure 2.1 A Typical Web Site Structure

A typical web site normally starts with one page called an index page (also known as entry page). A user visiting the web site would first see the index page and then pick and choose from the other pages linked. The following pages

normally would link to the index page where the hyperlinks was originated from it (the index page).

As a web site gets larger, this hierarchical tree structure becomes more difficult to navigate. Finding a particular page may become a tedious exercise of linking through unwanted pages. This is where the problems of disorientation and cognitive overload arise as discussed in Section 1.1.

2.2 TYPES OF WEB SITES

There are seven types of web sites [Embleton, 1998]. They are as the following:

1. Personal
2. Commercial
3. Archive/Reference
4. Current/News
5. Informational
6. Persuasive/ Propaganda
7. Educational

What differ these web sites are in terms of their purpose and number of contents.

A personal web site provides biographical information. It includes the person's name, address, date of birth, hobbies, their family, their photos, education level, and at times, resume as well. The number of pages in personal web site normally not exceeding 20 pages.

A commercial site provides information of their product or services. This serves as their main purpose of existence on the Internet. The number of contents and pages are normally not so many. One of the most successful commercial web site

is Amazon (<http://www.amazon.com>), a company selling books and other printed products online.

Some of the sites also exist purely for archive or reference purposes. This kind of web sites can be said as a content-rich web site as they would normally housed thousands and even millions number of documents. Digital libraries, virtual museums and archives web sites can be classified into this type of site.

Other sites contain up-to-date news and information, such as weather predictions or stock market quotes. They may not have so much number of documents but most of this kind of sites' contents are interactive in nature. This is due to the dynamic attributes of such information (i.e, stock prices and weather conditions).

Informational type of web sites provide information about hobbies, groups and etc. This type of web sites normally serve a specific audience. For example, the National Rifle Association of America (<http://nrahq.org/>) publishes information related to guns such as gun purchases, dealers, controls and firearms policy. The audience is normally the lawmakers and people associated in transacting firearms.

A persuasive site is one that pushes a particular viewpoint or dogma. These might be political or social in nature. One such example is the UMNO (<http://www.umno.org.my>), the ruling political party in Malaysia's web site. It gives information about the party's agenda and highlights their interest in the country's citizen welfare and future.

Finally, an educational web site exists to promote learning as their main purpose. All universities and colleges and even schools have their own web sites nowadays. The number of documents in this type of web sites vary but seldom exceed several thousands pages in number. If the number of contents offered are a lot, then it can classified as a content-rich web site.

2.3 CONTENT-RICH WEB SITES EXAMPLES

In this section, five content-rich web sites are reviewed. The web sites consist of one digital library, two digital archives and two digital museums. The first four web sites are from the United States of America and Australia. The fifth web site is from Malaysia – the National Archives of Malaysia web site.

Content-rich web site usually have a site map. Site map is a graphical representation of a web site's architecture [Rosenfeld and Morville, 1998]. This definition excludes table of contents and indexes that use graphic elements to enhance their aesthetic appeal. The hyperlink for a site map is normally put at the first page of the web site as shown in Figure 2.2. When the user click on the site map hyperlink, a new web page will appear. The site map shown in this page would normally represent the web site's architecture textually.

The web sites reviewed in the following sections cover the organization's brief information, collections and their contents' organization.

2.3.1 Library of Congress (<http://www.loc.gov>)

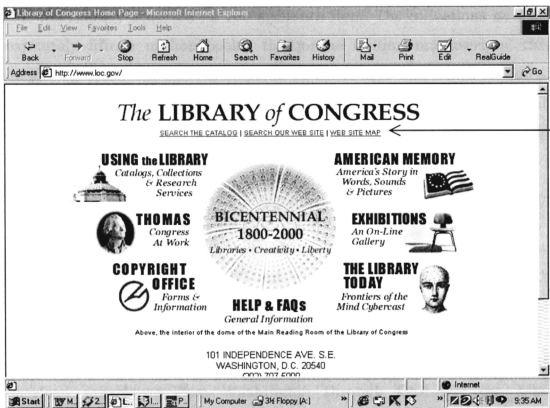


Figure 2.2 The Library of Congress Web Site

site map hyperlink

The Library of Congress occupies a unique place in American civilization. It was established as a legislative library in 1800 and grew into a national institution in the nineteenth century. The enormous size and variety of its collections make the Library of Congress the largest library in the world. Their collections are comprehensive in scope, including research materials in more than 450 languages and in many media. Each day about thirty-one thousand items arrive at the Library; approximately seven thousand of these items will become part of their permanent collections.

The Collections

In 1992, the Library acquired its 100 millionth item. The collections now include approximately fifteen million books, thirty-nine million manuscripts, thirteen million photographs, four million maps, more than three and a half million pieces of music, and more than half a million motion pictures. In addition, the Library holds newspapers, prints, posters, drawings, talking books, technical reports, videotapes and disks, computer programs, and other audio, visual, and print materials.

The Library collections are especially strong in American history, politics, and literature; music; geography; law and particularly foreign law; economics; genealogy and U.S. local history; U.S. public documents; publications of learned societies from around the world; the history of science; libraries and librarianship; and bibliography in all subjects. In addition to the personal papers of American presidents from Washington through Coolidge, the Library's manuscript holdings include the papers of eminent figures, mostly American, in government, the arts, and the sciences are also available.

The Site's Organization

The site's organization would normally be reflected back in their site map (as defined and discussed in the earlier section). One of the hyperlinks provided by this site is Web Site Map. It can be found in the first page of the web page. The Library of Congress implements their site map textually. When the Web Site Map hyperlink is clicked, Figure 2.3 will appear. By analyzing the site map, it is quite clear that its organization is done hierarchically.

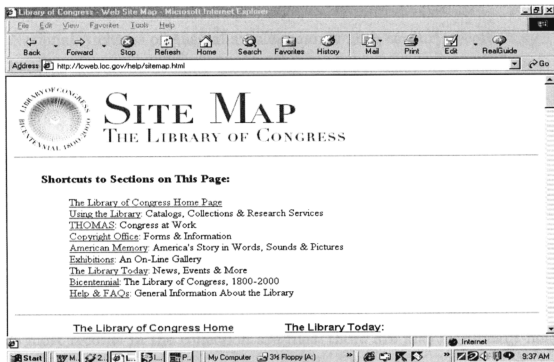


Figure 2.3 The Library of Congress Site Map

2.3.2 National Archives of Australia (<http://www.naa.gov.au/index.htm>)

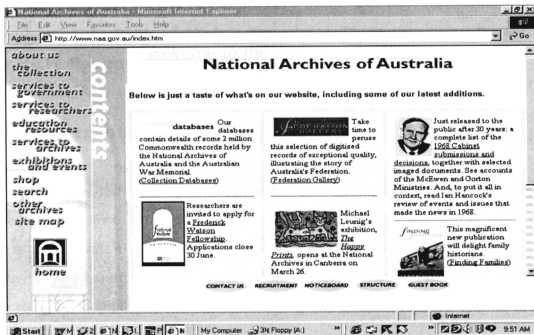


Figure 2.4 The National Archives of Australia Web Site

The National Archives of Australia holds mainly 20th century records created by the Federal Government since Federation in 1901. They also hold some 19th century records that were transferred from the colonies after Federation (e.g, records about defense, customs, patents, lighthouses, naturalization, shipping, and postal and telegraphic services).

The Collections

Spanning almost 200 years, the collection constitutes a vast and rich resource for the study of Australian history, Australian society and the Australian people. While most records in the collection are files, there are also significant holdings of photographs, posters, maps, architectural drawings, films, playscripts, musical scores and sound recordings.

The collection's main strength is material created since Federation in 1901 by agencies of the Commonwealth government, such as departments, royal commissions, statutory authorities, military units, security, intelligence and law enforcement agencies and diplomatic posts. These are complemented by substantial collections of nineteenth-century records.

The Site's Organization

As in the case of the Library of Congress, this site organization is also textually presented. The site map is hierarchical and arranged according the way it appears on the site's main page. One problem of the site map in this web site is the way it is presented. The text used is too small and it fills the whole right frame to be read without causing irritation to the users eyes.

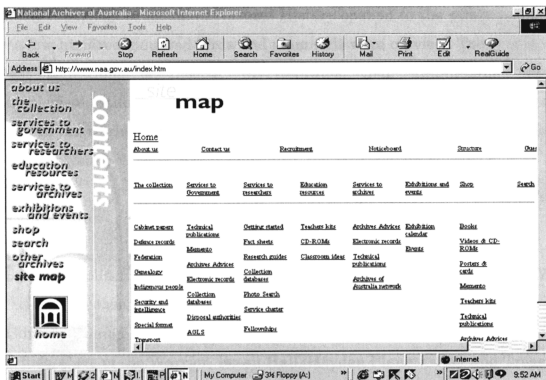


Figure 2.5 The National Archives of Australia Site Map

2.3.3 American Museum of Natural History (<http://www.amnh.org>)

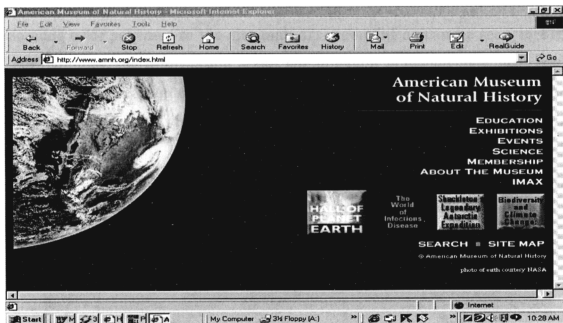


Figure 2.6 The American Museum of Natural History Web Site

When the American Museum of Natural History opened to the public on April 6, 1869, a few hundred mounted birds and mammals were on view. Today it is home to vast collections of insects, invertebrates, fish, amphibians, reptiles, anthropological artifacts, and more fossil mammals and dinosaurs than any other museum in the world. It has over 200 working scientists and welcomes millions of visitors each year.

The Collections

Among of the collections available includes the Hall of Planet Earth, The World of Infectious Disease, Shackleton's Legendary Antarctic Expedition, Bio-diversity and Climate Change. The Online Exhibit covers Endangered Animals, Tropical Butterflies, Cave of the Warrior, Fossil Hall, Hayden Planetarium and Treasures.

The Site's Organization

This web site has a very attractive site map. It also reflects the main page's iconic presentation of the site's content. However, instead of presenting it in plain text, it also adds color schemes to the site map arrangement as can be seen in Figure 2.7. For example, the information about the museum is placed in a table colored in pink, Exhibition contents in orange, Science and Research in cream, Education in green and Just for Kids in blue. The nature of the site map is indeed still hierarchical.

The coloring scheme used is a good approach in attracting users using the site while keeping them in context.

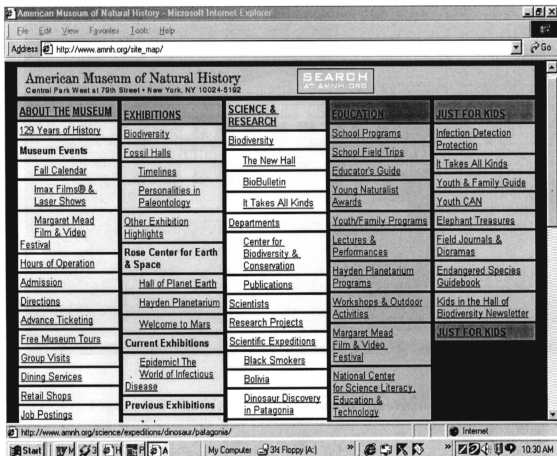


Figure 2.7 The American Museum of Natural History Site Map

2.3.4 National Museum of American Art (<http://nmaaryder.si.edu/home.html>)

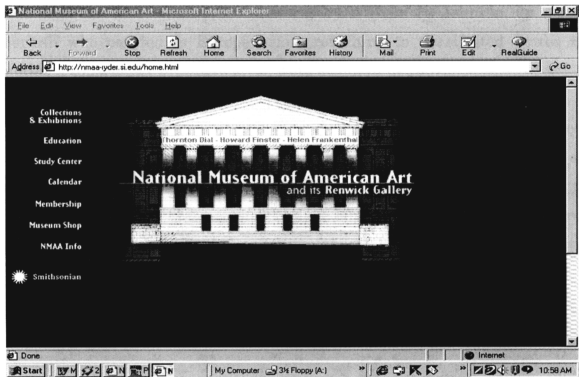


Figure 2.8 The National Museum of American Art Web Site

The National Museum of American Art is the home of the largest collection of American art in the world. Its holdings—over 37,500 works—represent the most inclusive collection of American art of any general museum today, reflecting the nation's ethnic, geographic, cultural, and religious diversity. The museum's roots go deep, representing three hundred years of American artistic achievement and paralleling since the collection began modestly in 1829.

The Collections

The collection began to grow as never before. Hundreds of paintings associated with New Deal projects as well as studies for post office murals were transferred to the museum. More than 1,300 artworks by William H. Johnson, an African American who painted in Europe before returning to the United States and adapting a consciously primitive style, were acquired in 1967.

Illustrative of the museum's diversity are other collections acquired in recent years: the extensive Herbert Waide Hemphill, Jr., collection of historic and contemporary folk art; major collections of both abstract and realist art from the 1920s–1950s; a trove of abstract works by the late Washington color-school artist Gene Davis; photography, including large-scale works; and a premiere collection of modern craft art.

The Site's Organization

The site map used in this site is hierarchical in its appearance. It also reflects back the original presentation of the table of contents provided in the left frame of the web page. The table of contents in this site is arranged according to its priority and done textually as well. The table of contents in the site is always visible. The purpose is to keep users in context while navigating the web site.

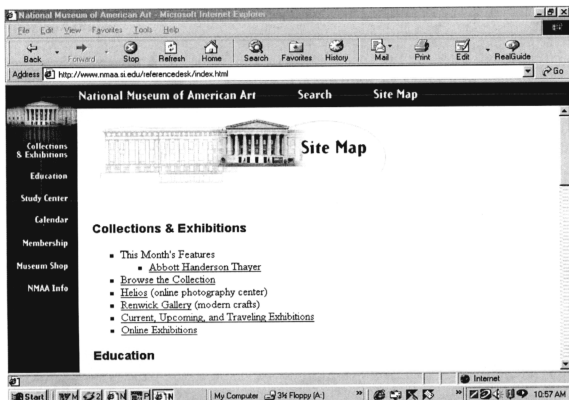


Figure 2.9 The National Museum of American Art Site Map

2.3.5 National Archives of Malaysia (<http://arkib.gov.my/>)

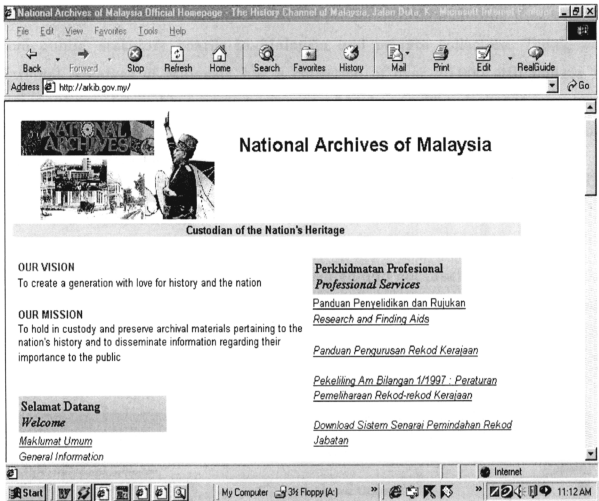


Figure 2.10 The National Archives of Malaysia Web Site

The first branch office of the National Archives now called the Johore-Malacca branch was established in Johore Bahru in 1972. This was followed by the setting up of various branch offices throughout the seventies and eighties including the Sarawak in Kuching (1976); the Sabah in Kota Kinabalu (1977), Kelantan Branch (1987) and lastly Ipoh (1993). With the formation of these branch offices, the National Archives is able to provide efficient services in the administration of public records throughout country.

The Collections

The National Archives of Malaysia collects public records including photographs from public, statutory boards in Malaysia and private collectors. Today, the National Archives of Malaysia has a collection of over 25,000 still images to present the social and economic landscape of Malaysia. The digitized version of the collections include The Legacy of Malay Letter, The Borneo Collection, Memorials, and The Electronic Classroom are available for Internet users.

The Site's Organization

The site's site map is also textually represented and not much different from the one offered on the main page. It is indeed an actual copy of the table of contents appeared on the main page. It is represented in hierarchical manner. The following figure shows the web site's site map.

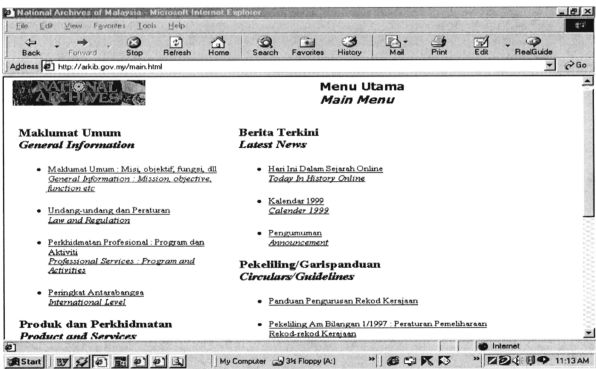


Figure 2.11 The National Archives of Malaysia Site Map

2.4 SUMMARY

The purpose of a web site always influenced of how they are structured. In the case of content-rich web sites, they are normally presented with a site map. This site map would naturally reflect back the site's contents and usually represented hierarchically.

As the site is growing in the number of its content, the site structure would become very big. Then, the problems of disorientation and cognitive overload would arise. This is where the visualization aspects are needed. The main emphasis of the following chapter is in the area of Information Visualization. The framework and some examples of information visualization interfaces will be discussed.